# DRAFT

May 12, 2006

# Database Management System Requirements Analysis Summary Document

## Lower Colorado River
## Multi-Species Conservation Program

**U.S. Department of the Interior**
**Bureau of Reclamation**

# Database Management System Requirements Analysis Summary Document

## Lower Colorado River Multi-Species Conservation Program

**DRAFT DOCUMENT**
**May 12, 2006**

**By:**
**LCR MSCP Data Management Team**
Theresa Olson, U.S. Bureau of Reclamation
Jeremy Dandron, U.S. Bureau of Reclamation
Christopher Flaccus, U.S. Geological Survey

# Table of Contents

# EXECUTIVE SUMMARY

The Lower Colorado River Multi-Species Conservation Program (LCR MSCP) is a multi-stakeholder Federal and non-Federal partnership responding to the need to balance the use of lower Colorado River water resources and the conservation of native species and their habitats in compliance with the Endangered Species Act. This is a long-term (50-year) plan to conserve at least 26 species along the LCR from Lake Mead to the Southerly International Boundary with Mexico through the implementation of a Habitat Conservation Plan. Most of the covered species are State and/or Federally listed special status species. The Bureau of Reclamation (Reclamation) is the entity responsible for implementing the LCR MSCP over the 50-year term of the program. A Steering Committee currently consisting of 54 entities has been formed, as described in the *LCR MSCP Funding and Management Agreement*, to provide input and oversight functions to support of LCR MSCP implementation.

In carrying out its mandate the LCR MSCP has and will continue to accumulate valuable legacy data sets from many sources. These data exist in a variety of formats and locations, which makes access to information difficult. A critical aspect of this conservation program is to organize, store and deliver relevant information in order to provide a sound base for the decision support process.

In October 2005, a data management team was formed to review existing environmental databases related to the LCR MSCP, to analyze the data and reporting needs of the LCR MSCP, and to provide recommendations to the Project Manager for the development of an interdisciplinary Data Management System (DMS). This document, a summary of the document entitled "Database Management Systems Requirements Analysis, Draft Document May 12, 2006," summarizes the data and requirements, recommends data management solutions, and provides a budget and schedule for DMS development. The full Requirements Analysis report includes this summary information, lays the foundation for the full DMS development, and provides additional information on Requirements of the full DMS system.

During the week of 12 December 2005, the data management team met with members of the LCR MSCP Administration group, LCR MSCP Steering Committee entities, and LCR MSCP scientific staff groups in order to evaluate problems, needs and opportunities of the program with respect to management of data. The problems, needs, and responsibilities determined from the interview process were grouped into key data related issues common to the various groups at the LCR MSCP. The data management team then formed four options for the DMS development.

The four options for the DMS presented in this document were developed based on meeting a broad range of needs and opportunities identified during the scoping process. Options 1, 2, and 4 concentrate on a multi-user, enterprise DMS that would fully achieve the mission, objectives, needs, and opportunities identified during interviews with LCR MSCP staff. Option 3 focuses on utilizing the current systems, which utilize separate databases, namely Microsoft (MS) Access and Excel, that Reclamation currently use in-house to organize data. This option meets basic database requirements and provides a low cost alternative to organizing data, it but does not meet all of the opportunities and needs identified during the scoping process.

# INTRODUCTION

The Lower Colorado River Multi-Species Conservation Program (LCR MSCP) is a multi-stakeholder Federal and non-Federal partnership responding to the need to balance the use of lower Colorado River (LCR) water resources and the conservation of native species and their habitats in compliance with the Endangered Species Act. This is a long-term (50-year) plan to conserve at least 26 species along the LCR from Lake Mead to the Southerly International Boundary with Mexico through the implementation of a Habitat Conservation Plan. Most of the covered species are State and/or Federally listed special status species. The Bureau of Reclamation (Reclamation) is the entity responsible for implementing the LCR MSCP over the 50-year term of the program. A Steering Committee currently consisting of 54 entities has been formed, as described in the *LCR MSCP Funding and Management Agreement*, to provide input and oversight functions to support of LCR MSCP implementation.

In carrying out its mandate, the LCR MSCP has and will continue to accumulate valuable legacy data sets from many sources. These data currently exist in a variety of formats and locations, which makes access to information difficult. Good data management is the key to maximizing the utility of the data, making it available to both managers and scientists to address adaptive management needs, as well as provide information to the LCR MSCP partners and general public. A well-designed data management system also improves the level of quality assurance and provides strong incentives to standardize and coordinate data collection. This all leads to transparent and timely reporting, making a more efficient and effective adaptive management program.

In October 2005, a data management team was formed to review existing environmental databases related to the LCR MSCP, to analyze the data and reporting needs of the LCR MSCP, and to provide recommendations for the development of an interdisciplinary Data Management System (DMS). This document, a summary of the document entitled "Database Management Systems Requirements Analysis, Draft Document May 12, 2006," summarizes the data and requirements, recommends data management solutions, and provides a budget and schedule for DMS development. The full Requirements Analysis report includes this summary information, lays the foundation for the full DMS development, and provides additional information on Requirements of the full DMS system.

# PURPOSE AND NEED

The purpose of this document is to outline the LCR MSCP data and requirements, to recommend data management solutions and to provide a budget and schedule for DMS development. The purpose of the DMS is to develop a data management system to consolidate, organize, document, store and distribute scientific information related to the LCR MSCP.

During the week of 12 December 2005, the data management team met with members of the LCR MSCP Administration, LCR MSCP Steering Committee, and LCR MSCP scientific staff groups in order to evaluate problems, needs and opportunities of the Program with respect to management of data. Scoping meetings were held with each group and the problems, needs, and opportunities were recorded on flipcharts. These were then consolidated and summarized into the following list:

1. A centralized database should be created to consolidate data, allowing integrated, long term analyses, and dynamic search ability with user friendly query tools to be performed to support adaptive management.
2. Many data collection, analysis and presentation software programs that are currently being used must be able to interface with any new data management system.
3. A more user friendly statistical analysis program is needed.
4. A unified, consistent file naming system is needed, along with managed file sharing.
5. Continuity with consistent data collection methodology should be enforced by a common database system, allowing for standardized format for forms and reports between projects.
6. Local onsite IT technical specialist support should be provided for training and clear instructions for new software and database technologies.
7. A dedicated data administrator is needed to manage consistency and storage of data.
8. Instantaneous field data download or FTP uploads are needed to incorporate data directly into database either while in the field or effortlessly upon returning from the field.
9. Library reference material database warehouse is needed for scientific papers and final project reports with keyword search features on documents related to individual studies.
10. Automated budget reporting, financial accounting and ability to generate requisitions, to incorporate with applications such as MS Access and MS Project would be valuable.
11. Project management system is needed for covered activities, including meeting management with calendars, agendas, notification, meeting minutes, attendee lists, meeting attachments, action items, motions and resolution tracking.
12. Geographical Information System (GIS) modeling capabilities will be greatly enhanced by developing a system that has strong integration between spatial GIS data and tabular data.

# LIST OF CURRENT DATA SETS

A list of current data sets was then compiled through individual interviews, and through the use of the current shared common drive. Each data set contained a few to several hundred files, depending on data being collected. Currently the data sets are contained on individual computers, with some of them being accessible on the shared common drive. The following is a list of the current data sets and their associated software applications:

| **Wildlife Group** | **Software Applications** |
|---|---|
| Neotropical Birds | |
| Area Search | Excel, GIS, Word |
| Point Counts | Excel, GIS, Word |
| MAPS – MAPSPROG, IBP, Bandmanager | Specific programs, Access, Excel |
| Winter, Migration banding | Specific programs, Access, Excel |
| Southwestern Willow Flycatcher | Access, Excel, GIS, External Contractor, D-Base |
| Yellow Billed Cuckoo | External Contractor |
| Marsh Birds | |
| Yuma Clapper Rail | Excel, Word, GIS |
| Small Mammals | Excel, Word, GIS |
| Bats | Anabat Files, GIS, External Contractor |
| Vegetation | |
| Total Vegetation Volume | Excel |
| Site Specific Habitat assessments | Excel, GIS |
| Photopoints | Photos-JPEG files |
| Microclimate and abiotic measurements | |
| HOBO's | Specific program, Excel, GIS |
| Soils | Excel, GIS |
| Water Quality | Excel, GIS |
| Yearly Technical Reports | Word, GIS, photos, PDF |
| Species Profiles | Word |
| Power Point Presentation | Power Point, PDF |
| Miscellaneous Pictures | JPEG files |
| Miscellaneous Studies (e.g. cowbird program) | Word, Excel, GIS, PDF |
| Aerial Photograph and Vegetation Type Mapping | GIS |

| **Restoration Group** | **Software Applications** |
|---|---|
| Water Quality/Soils data | Excel, GIS, Access |
| Architectural Drawings | AutoCad |
| Maps | GIS, PDF |
| Reports | Word, PDF |
| Designs | AutoCad, GIS, PDF |
| Photos | JPEG format |
| Presentations | Power Point, PDF, Word |
| Project Management | Microsoft Project, Word, PDF |
| Chronicles | Word |
| Site Selection Criteria | Word |

| **Administration Group** | **Software Applications** |
|---|---|
| Workplans | MS Word |
| Payroll | FPPS, T&A in Excel or Word, TAAS, FFS |
| Financial Management | FERS, Excel, Specific programs |
| Cost Authorities | Excel |
| Accruals | Excel |
| Budgets | Excel, Word |
| Accomplishment reports/monthlies | Excel |
| Acquisitions | Excel, Word |
| Staff Costs | Excel, Word |
| Misc. Spreadsheets | Excel, Word |
| Misc. Reports | Word |
| Calendar | Word, Groupwise |
| Project Schedules and updates | Word, Microsoft Project, Groupwise |
| Communications Outreach | Power Point, JPEG Photos, Word, Groupwise, Internet |
| Correspondence | Word |
| Photos | JPEG format |
| Steering Committee Management | Word, Power Point, Excel, Internet |

**Fisheries Group**

The fisheries group data is housed in a separate database at Arizona State University. This database includes data for mark-recapture and other studies related to bonytail, humpback chub, flannelmouth sucker, and razorback sucker for locations along the LCR from Glen Canyon Dam to the U.S.-Mexico International Border. These databases are currently maintained in Excel and Access formats. Additional data is stored in Reclamation's shared drive and on individual computers including: all reports are completed in Word, photos are housed in JPEG Format, larva collection is maintained in Excel, hydro-lab and netting data are in Access, and backwater elevations are in Excel.

# OPTIONS

The following four options are proposed for system development. All of these options would seek to improve the data management and reporting capabilities of the LCR MSCP staff by organizing data collected by the agency based on projects or subjects outlined in the LCR MSCP Work Plans.

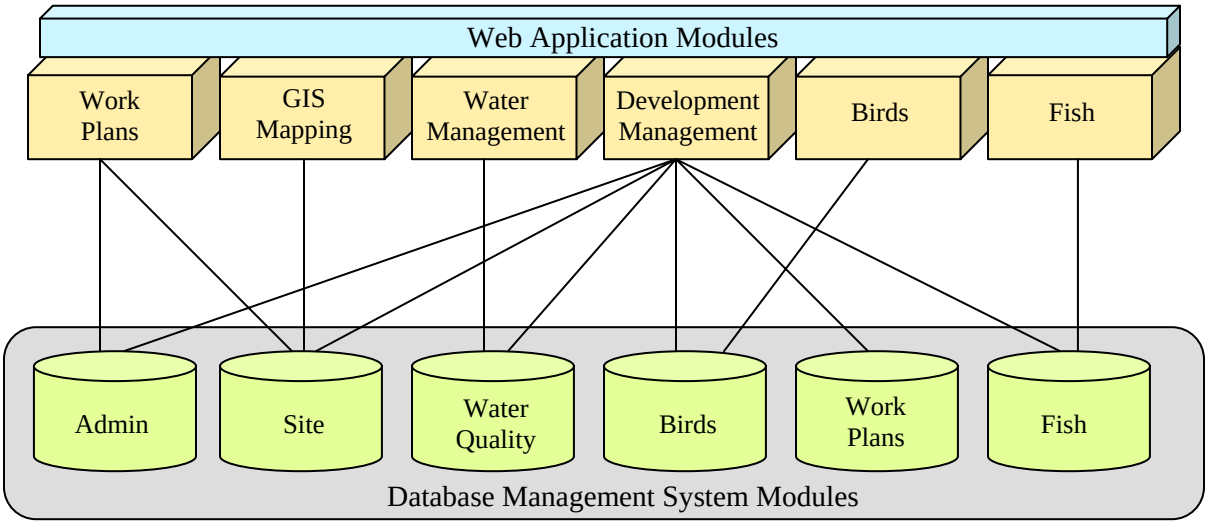
Options 1, 2, and 4 concentrate on a multi-user, enterprise DMS that would fully achieve the mission, objectives, needs, and opportunities identified during interviews with LCR MSCP staff. Option 3 focuses on utilizing the current systems, which utilize separate databases, namely Microsoft (MS) Access and Excel, that Reclamation currently use in-house to organize data. This option meets basic database requirements and provides a low cost alternative to organizing data, but it does not meet all of the opportunities and needs identified during the scoping process. Other options that were considered, but rejected, include: 1) obtaining the full Southern Nevada Water Authority (SNWA) database "as is" with no further refinement, rejected because of security reasons, coding concerns, and not meeting the full needs of the LCR MSCP biologists; 2) other data management systems currently on the market rejected because most of these were developed for business and not for research and development.

## Option 1 – Sole Solutions Integration

In Option 1, the LCR MSCP would seek to develop a DMS that would be a comprehensive, integrated database that would provide a secure, reliable, and productive platform for making data more manageable. The DMS would be a relational database management system (RDBMS) that houses all data in one database to include all LCR MSCP data and information including scientific data, financial, and meeting management. The DMS would utilize a web interface for easy access for all users including technical, management, and LCR MSCP Steering Committee members, along with the public, at different security levels.

This option would improve the data management and reporting capabilities of the LCR MSCP staff by organizing data collected by the agency into individual modules based on projects or subjects outlined in the LCR MSCP Work Plans, all stored in one large database. A database module is a grouping of like information into a management unit, or a logical grouping of similar information around existing projects within the LCR MSCP (Figure 1). For example, water quality data would be grouped into one module that would be used to manage that information. These modules would be used to design relational tables to store information, and data would be imported into those database structures. Finally, applications, or programs, would be written to aid users in extracting and analyzing information collected, making the system user friendly.

Figure 1: Example of the Web-Application Module and Database Module Architecture




The DMS would be completed in a phased process as outlined below under *PHASES OF DEVELOPMENT*. The DMS would be modeled after SNWA's database system, and be completed by specific in-house experts. Skills needed for this expertise are listed below under *STAFFING SKILL REQUIRMENTS*. Some programming tasks could be completed by contractors, but the database manager would be a Reclamation employee. Existing Lower Colorado Regional Office Information Technology (LCRO-IT) and Lower Colorado Regional Office GIS (LCRO-GIS) staff would be utilized to their fullest extent and whenever possible. With such an aggressive approach, a mature database system could be developed within three years, and all institutional knowledge of the system gained during the development phase would be retained by LCR MSCP staff members.

LCR MSCP staff would utilize much of the SNWA database and application structure as models for system development, thereby "jump-starting" the LCR MSCP system development and applying lessons learned by the SNWA system developers and users. An initial pilot project would be used to test the applicability of certain SNWA modules to the LCR MSCP.

Beyond the development phase of the DMS, in years 4-8, emphasis would change to maintenance and fine tuning the database and applications developed during the first three years. During years 9-20 basic system maintenance and data management would continue.

**Total Cost Summary for Option 1:**

The estimated costs for this option are based on the following assumptions:

- Three Full Time Equivalents (FTE) to include skills for a Software Application Developer, a Database Developer, and a GIS developer, at an average cost of $80/hour for GS-11/12 level employees.
- Two contract programmers at a rate of $80/hr, not counting travel expenses.
- All computer hardware and software to be purchased and maintained by LCR MSCP, LCRO-GIS and LCRO-IT staff.
- Functional system would be completed within 3 years.

- After 3 years, contract programmers would not be needed.
- After 8 years, database may be maintained by 1-2 FTE's.
- Three year replacement on computer hardware for maintenance period.

Total Cost for Option 1:

| | | |
|---|---|---|
| Development years 1-3 ($980,722/year) | $2,942,167 | |
| Maintenance years 4-8 ($520,039/year) | $2,600,195 | |
| Total 8 year cost | | $5,542,362 |
| Maintenance years 9-50 ($200,000-400,000/year) | 8.4-16.8 Million | |

**Total Maximum Fifty Year Costs = $22,342,362**
**Total Minimum Fifty Year Costs = $13,942,362**

**Benefits**

- Final DMS would be user-friendly and meet the specific needs of the LCR MSCP.
- All data (scientific, administration, GIS, project management, etc) would be stored in one large relational database, making it efficient to consolidate, retrieve, and analyze large data sets across disciplines over a 50-year period.
- Application programs would be created to allow the database to be searchable in multiple ways using user friendly query tools.
- A common database system would provide an effective tool for decision support and adaptive management of resources.
- The relational database would maximize the utility of the data, making it readily available to managers, scientific staff, and the general public.
- Monitoring results and restoration activities could be reviewed efficiently for quality and effectively shared for transparent and timely reporting.
- Communication would be enhanced by providing web interfaces for access, input, and extraction of data. Upload of data from outside sources, such as contractors, would be more efficient and would be able to be quality controlled before placing into the database for ready access and analysis.
- This relational database would provide continuity of data collection and analysis through the use of quality controlled, standardized input screens that would have automatic quality assurance and quality control built into the system. There will be checks and balances built into the system for input fields. For example, screens would only allow certain codes that would belong in that input field; therefore a user would not be able to corrupt the system by inputting wrong codes.
- In-house, local database management and development assures that specific needs of LCR MSCP would be met through close interaction between scientists and database developers.
- Institutional knowledge gained during the development cycle would be retained by the LCR MSCP in the long term.
- Big-picture view of the system would be established early in the development cycle. A modular approach to system development would assure that future needs of the LCR MSCP would be efficiently integrated.
- Increase in production performance may be realized through saved time in entering and retrieval of data and modeling of results. For example, in Option 1 users can automatically upload data through the use of applications, whereas if this is not a priority in the

development of Option 2, the data would be manually entered, versus direct download. Option 3 would also be manual input of data.
- High probability of early success would be achieved by utilizing design and programming of SNWA system.

**Drawbacks**
- Increases the number of FTE positions for LCR MSCP.
- Larger up-front costs for LCR MSCP.
- Places significant responsibility and accountability on LCR MSCP to find the right personnel to make the project successful.

## Option 2 – Prioritized Technical Module Development

Option 2 is similar to Option 1 in that the DMS would still be a comprehensive, relational database, but the database would be constructed on a module-by module basis, with development prioritized based on need, functionality, and cost, with scientific data requirements receiving top priority. This option would improve the data management and reporting capabilities of the LCR MSCP staff, by organizing data collected by the agency into individual modules as defined in Option 1, but the modules would be prioritized by a committee composed of LCR MSCP staff. This option would seek to lower expenses by reducing the number of full time staff dedicated to the project, focus on developing the technical and scientific components of the database, and extend the time over which development occurs to five years. Up front system planning and design would be minimized. This Option would also utilize the Phase approach as discussed below in the *PHASES OF DEVELOPMENT* section.

Modules having top priority would include:

- Monitoring and Research: including data input and retrieval for the 26 LCR MSCP covered species and their habitats focusing on birds, bats, small mammals, vegetation, soil, water quality, and microclimate parameters.

- Habitat Development and Management: project management, water management, land development and site management.

- Incorporation of GIS and Mapping and linking to the Monitoring and Research and Habitat Development and Management modules.

- Systems that are low cost and/or can be migrated from SNWA and Reclamation's Phoenix Area Office, including meeting management and library functions.

Lower priority modules would include administrative tasks such as training, workplans, financial, and contracts. A fisheries module would also be considered lower priority because a functioning database is already in existence (see *Current List of Data Sets* section). As each module is finalized it would become fully functional and usable by LCR MSCP staff. With this approach, all high priority modules, and many of the lower priority ones could be fully developed within a five year time frame. Beyond five years, the system would enter a

maintenance mode; new module development would be minimized, and emphasis would be placed on data maintenance and analyses.

This option constitutes a pay-as-you-go approach to system development. Two permanent staff (FTE's) would be hired by Reclamation in order to coordinate and perform system development. Permanent employees dedicated to the project would assure close communication with other LCR MSCP staff, local "ownership" of the process, and retention of institutional knowledge gained during system development. Existing LCRO-IT and LCRO-GIS staff would be utilized to the maximum extent whenever possible, but all other development tasks, including application programming, would be contracted on a project-by-project basis. Whenever possible, student contractors and interns would also be utilized for routine data cleaning, coding and importing tasks. The two additional staff would be retained after the five year development stage, but it is assumed that the need for contracting development work would diminish beyond year 5. This approach to application development gives flexibility in that budgeted money does not need to be spent unless a specific need exists.

In this option, a partnership with SNWA would be beneficial to jump-start the development process. Some of the database structure and applications developed by SNWA would be implemented on the LCR MSCP database server, increasing the likelihood of an early success in the development process. A pilot project would be undertaken first, in which data from a single LCR MSCP discipline would be used to tailor the SNWA database and application structure to fit the specific needs of the LCR MSCP.

As database modules are developed and put into production, they would continue to require maintenance, data loading, consistency checks and minor modification. So, while the character of the FTE positions identified would change somewhat through time, the staffing needs of the DMS beyond the five year development cycle would likely remain constant. At the end of the five year development cycle it is assumed that the need for contract programmer/analyst would be minimal, perhaps limited to minor modifications of existing code. Furthermore, a student contractor may not be necessary to help maintain the system. Therefore, values given for these contract positions in years 6 and beyond represent estimates of potential needs, and are subject to available funds.

**Total Cost Summary for Option 2:**

The estimated costs for this option are based on the following assumptions:

- Two FTE's to include skills for a Database Developer and a GIS developer, at an average cost of $80/hour for GS-11/12 level government employees.
- A single contract programmer with a rate of $80/hr, not counting travel expenses.
- A single student contractor with a rate of $14/hr, not counting expenses.
- All computer hardware and software to be purchased and maintained by LCR MSCP, LCRO-GIS and LCRO-IT staff.
- Functional system would be completed within 5 years.
- Maximum 5-8 year maintenance funding includes $30,000/yr for a contract programmer/analyst.
- Maximum 5-8 year maintenance funding one student.
- Minimum 5-8 year maintenance funding $0/yr for a contract programmer/analyst.

- Minimum 5-8 year maintenance funding no student.
- Maximum 9-50 year maintenance based on 2 FTE's and upgrade of hardware on a 3 year cycle.
- Minimum 9-50 year maintenance base on 1 FTE and upgrade of hardware on a 3 year cycle.

Total Cost for Option 2:

| | | |
|---|---|---|
| Development years 1-5 ($479,513/year) | $2,397,565 | |
| Maintenance years 6-8 ($345,265-$403,265/year) | $1,035,797-$1,209,797 | |
| Total 8 year Costs | | $3,433,362-$3,607,362 |
| Maintenance years 9-50 ($200,000-$400,000/year) | 8.4-16.8 Million | |

**Total Maximum Fifty Year Costs = $20,407,362**
**Total Minimum Fifty Year Costs = $11,833,362**

**Benefits**

- Most of the summarized needs above would be met, and most of the benefits from Option 1 would be realized.
- Once modules are fully developed this will be a powerful tool for comparison of scientific data within and across species and habitat creation areas. GIS models would be enhanced by the ability to analyze tabular data with GIS data.
- Once modules are fully developed, data will be easily accessed and be easily analyzed through the applications produced.
- Each separate module will be ready to use after its development.
- Lower initial and recurring costs and fewer FTE positions than Option 1.
- High probability of early success utilizing design and programming of SNWA system.
- Pay-as-you-go process allows LCR MSCP to re-evaluate development on a module-by-module basis.
- Partnership brings expertise and experience of SNWA to the ecological database system development process.
- Retention of institutional knowledge gained during development is good, particularly when long term benefits of partnership are considered.

**Drawbacks**

- Ability to compile and analyze data across species and conservation areas would be delayed and data comparison between modules would lag until scientific modules are complete.
- Because all modules would not be developed at one time, big picture view of database may be initially lacking. Option 2 would have limited planning before going into the production. Although priorities would initially be identified, these may be changed part way through the process, thus coding may be affected, or may affect other modules that were previously developed.

- Because of prioritization of modules, some areas such as administration tasks, finance, and web interfaces may either be delayed or never included in one large database, thus some aspects of the LCR MSCP program may still be disconnected from the larger database.
- System development timeline may be extended due to the pay-as-you go system.
- Adapting the system to changing needs in the future may be more difficult than Option 1. Extending the system to accommodate new LCR MSCP projects may be more difficult in Option 2, due to legacy design and programming constraints. This is not the case in Option 1 because work flow would be analyzed prior to development of the system, whereas in Option 2, work flow would be analyzed on demand as system is built.
- Debugging existing code may be more difficult in Option 2 versus Option 1, because in Option 2, with the pay-as-you-go plan, the system would not be fully planned in advance, therefore affects of previously written code may affect code written later, and visa versa. In Option 1, full system development would be thoroughly planned; therefore codes interfering with each other would be reduced.
- Longer development time than Option 1.

## Option 3 - Existing System Enhancement

This option utilizes current software and applications such as, MS Access and Excel, with little or no web-application functionality and no centralization of data into one large database. A dedicated data manager would be hired to organize legacy data, develop databases using MS Access, and provide support to existing LCR MSCP scientific staff. The existing data shared common drive (Y: drive) would be restructured to organize data and to provide a location for MS Access databases that are developed to house the data. This option would rely heavily on users to maintain data for which they are responsible. Security would be provided using MS Active Directory Services. Reporting would also largely be the responsibility of users, with the help of the data manager. The environment proposed under this option would not allow for enterprise architecture, so there would be no single database repository that would allow for analysis of multi-discipline data sets (e.g. bird data would not be readily linked to vegetation data). Concurrent users would not be able to have access to a shared MS Access database at the same time. Due to this limitation, the data manager would set and maintain the standards and protocols for storing data within the centralized data folder. Data development would be ongoing, with priorities set by the data manager.

All computer hardware and software maintenance and support would be provided by the LCRO-IT staff. Data maintenance would be provided by the data manager and all data backup support would be provided by the LCRO-IT. All current GIS support would continue to be provided by the LCRO-GIS staff.

This type of data storage and management results in scope creep, slow production, and data loss. Control over data production and integrity is the key to the implementation of the Adaptive Management program, upon which the success of the LCR MSCP depends. This option does not allow for efficient quality control. A data manager will help decrease loss of data but at a nominal rate.

**Total Cost Summary for Option 3:**

Maximum and minimum costs for this option are based on the following assumptions:

- One FTE to serve as a Data Manager, at an average cost of $80/hour for a GS-11/12 level government employee.
- No additional contract positions.
- No purchase of addition database software.
- All computer hardware and software to be purchased and maintained by LCR MSCP, LCRO-GIS, and LCRO-IT staff.
- System development would be on-going.
- Three year replacement on computer hardware.

Total Cost for Option 3:

| | | |
|---|---|---|
| Development years 1-5 ($167,920/year) | $839,600 | |
| Maintenance years 6-8 ($229,667/year) | $689,000 | |
| Total 8 year Costs | | $1,528,600 |
| Maintenance years 9-50 ($180,000/year) | $7,560,000 | |
| Total 50 year Costs | | $9,088,600 |

**Benefits**

- Most users are familiar with using the current programs of Access and Excel. Data would not need to be transferred into a new database system.
- This option would allow for the consolidation of data sets in similar formats onto one shared drive, but individual files would be maintained.
- Data manager would organize current and future data, and would help with standardized input of data.
- Fewer FTE positions than either Option 1 or 2.
- Lower annual costs than either Option 1 or 2.
- Development would focus on immediate needs of the LCR MSCP.

**Drawbacks**

- Does not meet most of the summarized needs.
- Coordination and consolidation of data would be the responsibility of each individual researcher to gather various and numerous files for analysis for each study or project.
- No single data repository for effective and efficient analyses of data across species or across conservation areas.
- Data loss and/or corruption a very real and probable possibility, because of lack of automated quality assurance/quality control.
- Institutional knowledge of data may be lost due to staff turnover.
- Metadata standards ensuring quality assurance/quality control may be difficult to enforce.
- Long-term vision and planning may be lacking.
- Current software programs such as MS Access and Excel have limitations on maximum file size and record numbers.
- Would result in a disjointed system design that is less scaleable.
- Needs of LCR MSCP scientists may not be met due to inefficiencies and inconsistencies in data input and storage.

## Option 4 – Outsourced System Development

This option would create a DMS by utilizing a large-scale contract with a consulting firm to fully develop the system and populate the database. Application design, development, coding, implementation, and maintenance of the system would all be contracted. A principle advantage to this approach is that database expertise is purchased, and responsibility for successful DMS completion would lie with the consulting company chosen to develop the system. Furthermore, the system could be designed to meet all of the mission goals within three years of contract award, and no additional FTE positions would be required until development is complete. Disadvantages include a high initial cost, and potentially a long term reliance on the consulting firm for future system modifications and enhancements. Little to no institutional knowledge of the development process would be retained within the LCR MSCP. Effective communication between customer and contractor would be critical to an effective and useful system.

**Total Cost Summary for Option 4:**

Costs for this option are based on the following assumptions:

- Project management would be provided by existing LCR MSCP staff during development.
- One FTE to act as a Data Manager when development is complete, at an average cost of $80/hour for a GS-11/12 level government employee (this could also be contracted).
- All development expertise would be provided by the contracting company at an average hourly rate of $150/hr for three years.
- All computer hardware and software would be purchased and maintained by LCR MSCP, LCRO-GIS and LCRO-IT staff.
- Three year replacement on computer hardware for maintenance period.
- Detailed requirements analysis would be performed by contracting company at approximate cost of $200,000.
- Functional system would be completed within 3 years.
- After 3 years, system maintenance would be provided by contracting company.
- After 8 years, system maintenance would be provided by contracting company.

Total Cost for Option 4:

| | | |
|---|---|---|
| Development years 1-3 ($1,758,333/yr) | $5,275,000 | |
| Maintenance years 4-8 ($938,000/yr) | $4,690,000 | |
| Total 8 year costs | | $9,965,000 |
| Maintenance years 9-50 | $200,000-400,000/year | $8.4-16.8 Million |

**Total Maximum Fifty Year Costs = $26,765,000**
**Total Minimum Fifty Year Costs = $18,365,000**

**Benefits**

- Database system could be sized to fit total needs of staff.
- If the contractor would be tasked with an enterprise, relational database system, as in Options 1 and 2, the benefits listed in those Options would apply.
- Database system would be designed to meet all of the mission goals within three years of contract award.
- Places the responsibility for successful project completion on an outside database development consulting firm.
- No additional FTE positions during development phase.

**Drawbacks**

- Long term reliance on consulting firm for future system modifications and enhancements and system maintenance.
- May not get exactly what LCR MSCP staff needs in a system.
- Little to no institutional knowledge of the development process is retained by LCR MSCP.
- Long and potentially tedious contracting process.
- Little day to day interaction between LCR MSCP staff and system developer
- Highest development costs of the four options.
- Highest overhead/maintenance cost of the four options.
- Currently, there are no commercially available software products available for this type of ecological database; therefore it would have to be developed as a custom application, with potential proprietary issues.

# PHASES OF DEVELOPMENT

Options 1, 2, and 4 for database development were based on developing systems that would meet the needs and opportunities identified during the scoping process. Option 1 and potentially Option 4 would have a full system design mapped prior to development, whereas Option 2 would have a limited system design before going into the phased portion of the development, because of prioritization, and the pay-as-you-go plan. A three-phase development effort is proposed for Options 1, 2, and 4. The planned phases are:

| | |
|---|---|
| **Phase I** | Pilot |
| **Phase II** | Production |
| **Phase III** | Ongoing Deployment of Data sets |

**Phase I - Pilot**

The use of a pilot phase would allow for an incremental architectural DMS design and development process using a series of iterative steps or milestones. For the Pilot, a single LCR MSCP project would be selected to produce an initial database module design. Individual data sets comprising the selected project would be modeled, and database tables created and data imported. Applications would then be written to manage and access the data. Once pilot development was complete, feedback from LCR MSCP staff and stakeholders would be integrated to assure that the long term needs of the LCR MSCP would be met by the design.

The goal of the Pilot would be to produce the following products and features:

- Installation and configuration of computer hardware and RDBMS, ArcSDE, and ArcIMS software to be used in DMS development.
- Creation of RDBMS tables to store tabular data related to the selected project.
- Loading of RDBMS tables with real data from Pilot Project.
- Creation of Pilot Project GIS layers in ArcSDE.
- Creation of user accounts and security model granting appropriate access to related data.
- Creation of Web applications to facilitate loading and management of data associated with the Project.
- Creation of Web applications to access data sets and reports associated with the Project.

On completion of the Pilot Project, LCR MSCP staff and stakeholders would evaluate the results.

**Phase II – Production**

Phase II of the development effort would seek to put the Pilot application into production. This would involve making architectural changes to the database and applications based on user feedback and technical knowledge learned during the Pilot Phase. Phase II would finalize any code developed during the Pilot into a production quality application on the LCR MSCP intranet. At the end this phase, pilot data sets would be fully accessible using Web and GIS tools, and specific analysis needs of each user group, identified during the Pilot, would be met.

**Phase III – Ongoing Deployment of Data Sets**

Phase III of the development effort would seek to add remaining data sets into the database, and would build on the product developed during Phase II. Phase III deployment would be driven by priorities as defined by the LCR MSCP Program Manager and staff. Datasets comprising remaining projects identified in the Work Plans would be evaluated and modeled. Once a data set has been analyzed to ensure consistency and value, it would be added to the DMS, and applications would be written giving access and manageability to the appropriate users.

In summary, development of the DMS should be according to the following rough schedule:

- Determine and/or hire a database manager early on in the development process.
- Coordinate with Denver and Regional USBR Offices to obtain hardware and software systems.
- Coordinate with LCRO-IT and LCRO-GIS groups to determine staffing opportunities and needs.
- Hire additional technical staff to work on systems development. With LCR MSCP technical staff involvement early in the project, knowledge of system functionality developed by contractors will then be retained by the organization.
- Detail all components and modules needed for the system during interviews with LCR MSCP staff, and share that information with the customers.
- Obtain buy-in from LCR MSCP staff with respect to functionality and priorities before development of each component or module
- Priorities for system development include:
    1. Develop and implement a Pilot project of limited scope.
    2. Further interview staff to refine development priorities and schedules.
    3. Develop a user friendly GIS interface.
    4. Develop an annual report/Work Plan interface. This will lay the foundation of how the projects are to be set up in the database, as all projects will be Work Plan related.
    5. Develop a module for metadata.
    6. Develop specific data forms and project specific modules for Wildlife group based on prioritized needs.
    7. Develop specific modules for Restoration group based on prioritized needs.
    8. Develop specific modules for Fishery group based on prioritized needs.
    9. Provide for miscellaneous projects based on prioritized needs.
    10. Develop a Meeting Management system for steering committee and in-house staff.
    11. Develop interfaces for outside entities, including upload and download functionality.
    12. Integrate legacy data (prior to implementation of LCR MSCP, such as past BA/BO requirements that have an impact on current projects), as need arises, as appropriate. Examples include Southwestern Willow Flycatcher program data, previous vegetation data, previous Yuma clapper rail data, etc.

# STAFFING REQUIREMENTS

The following is a summary list of skills sets required to both develop and support the DMS as described in Options 1 and 2 of this document. It is possible to find one or more individuals that may fulfill all of these skill sets. Furthermore, existing personnel within LCRO-IT and LCRO-GIS may possess combinations of these skills sets, and could be utilized during the development phase of this project.

It is likely that System and Database administration support would be provided by existing LCRO-IT staff at Reclamation. It is also likely that the LCRO-IT and LCRO-GIS groups would provide some application development expertise, notably in the area of Internet Map Service (IMS) development and maintenance. It is therefore recommended that at least two additional full time staff be dedicated to the DMS development and support. One of these would be primarily a database developer/data manager, and one an application developer/manager. The combination of LCR MSCP and existing staff dedicated to this project would include skills in the following software and hardware:

**Computer System Administration -** responsible for general administration of the server computer dedicated to running both RDBMS and ArcSDE software products. Tasks include operating system installation and upgrades, and backup and recovery of system and database files. It is likely that existing LCRO-IT group can supply this expertise for the LCR MSCP server computer.

**Database Administration (DBA) -** responsible for installation, configuration, tuning, and backup and recovery of the database software (either MS SQL Server 2005 or Oracle ver. 10.20). It is likely that with some additional training, existing LCRO-IT staff can provide this support for LCR MSCP.

**Database Development -** responsible for developing the initial database design, naming conventions and security design, and creating tables, constraints and relationships between entities for the entire database. A successful developer would have a deep understanding of the science requirements behind the data that will go into the database.

**GIS Developer/Analyst -** responsible for the spatial aspects of system development, ensuring that all spatial data is stored within the DBMS in an efficient manner and that spatial information can be extracted and displayed using current Web technologies. Other responsibilities could include: ArcSDE installation, configuration and management, GIS support including knowledge in ArcSDE data storage and techniques, and lending support to LCR MSCP staff for various GIS modeling tasks.

**Data/Information Management and Support -** knowledgeable in all data collection activities for the LCR MSCP, the day to day operation of RDBMS software, maintenance of user accounts, data import and export, data requests and SQL query development and support, etc. Responsibilities would also include long term maintenance of the data resources at LCR MSCP.

**Application Development and Support** - responsible for web development applications such as VB.NET, ASP.NET and MS IIS and for creation of custom applications used for extracting information from the database and presenting it to the staff, Steering Committee, and general public.

**General PC Application Support -** skilled in all of the standard applications used by the LCR MSCP (MS Access, Excel, SAS, SYSTAT etc.), would troubleshoot specific issues with PC applications, and would provide specific training for LCR MSCP staff.

# APPENDIX I

# List of Acronyms

| | |
|---|---|
| ANSI | American National Standards Institute |
| ArcIMS | Arc Internet Map Server - Software from ESRI used to display maps and other GIS related information to the Internet |
| ArcSDE | Arc Spatial Data Engine – A GIS software product from ESRI that allows the storage of spatial and tabular data in mainstream database management system products such as MS SQL Server or Oracle |
| AutoCAD | A popular Computer Aided Drafting (CAD) program from AutoDesk, Inc., commonly used in architectural and engineering site plan preparation |
| BA/BO | Biological Assessment/Biological Opinion. |
| Bandmanger | Computer program used throughout the USA and Canada for reporting banding data to Bird Banding Offices |
| DBA | Database Administrator |
| DMS | Data Management System |
| ESRI | Environmental Systems Research Institute – A company that develops and markets GIS software |
| FTE | |
| FTP Site | File Transfer Protocol Site – A computer site that allows users to easy transfer of large data sets using a standard software independent process |
| GIS | Geographical Information System - Computer software system allows for the storage, display and analysis of spatial information |
| HOBO© | Data logger instrument used to measure temperature and relative humidity |
| IMS | Internet Map Service – A computer server application by ESRI, used to display GIS related maps and information on the Web |
| IT | Information Technology |
| JPEG | Joint Photographic Experts Group – a standard digital file format used for drawing and photographic images (.jpg) |
| LCR | Lower Colorado River |
| LCR MSCP | Lower Colorado River Multi-Species Conservation Program |
| LCRO-IT | Reclamation's Lower Colorado River Information Technology staff |
| LCRO-GIS | Lower Colorado Regional Office GIS staff |
| MAPSPROG | Computer program used to manage bird banding data for MAPS |
| MAPS | Monitoring Avian Productivity and Survivorship |
| MS SQL Server | A Relational Database Management System (RDBMS) marketed by the Microsoft Corporation |

| | |
|---|---|
| Oracle | A Relational Database Management System (RDBMS) marketed by the Oracle Corporation |
| PDF | Portable Document Format – A standard document format developed by Adobe Corporation designed to facilitate document sharing between parties |
| RDBMS | Relational Database Management System |
| Reclamation | U.S. Bureau of Reclamation |
| SAS | Statistical data analysis software from the SAS Institute Corporation |
| SNWA | Southern Nevada Water Authority |
| SQL | Structured Query Language – An ANSI language and syntax specification for communicating with a database management system |
| SYSTAT | A statistical analysis software from Systat Software, Inc. |